AD_____

Award Number DAMD17-96-1-6145

TITLE: Breast Ultrasound Computer-Aided Diagnosis Approach to Improving Specificity and Decreasing Observer Variability

PRINCIPAL INVESTIGATOR: Joseph Lo, Ph.D.

CONTRACTING ORGANIZATION: Duke University Medical Center
Durham, North Carolina 27710

REPORT DATE: January 1999

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE January 1999	3. REPORT TYPE AND DATES COVERED Annual (1 Jan 98 - 31 Dec 98)	
4. TITLE AND SUBTITLE Breast Ultrasound Computer-Aided Diagnosis Approach to Improving Specificity and Decreasing Observer Variability		**5. FUNDING NUMBERS** DAMD17-96-1-6145	
6. AUTHOR(S) Joseph Lo, Ph.D.			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Duke University Medical Center Durham, North Carolina 27710		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**	

19990820 017

11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited	**12b. DISTRIBUTION CODE**

13. ABSTRACT *(Maximum 200 words)*

The purpose of this grant is to construct an artificial neural network (ANN) to assist radiologists in differentiating benign from malignant solid breast lesions. The project was delayed due to a change in the principal investigator (PI) during this past, second year. As a result of those delays, a no-cost extension for a third year has been requested and approved. Much of the work originally planned for the second year has been refocused and rescheduled for the upcoming third year. These include more data taking and some final analyses to evaluate the performance of the ANN computer models. Continuing the work from the first year, the ultrasound (US) examinations and mammograms of 100 solid breast lesions that subsequently underwent histologic confirmation were evaluated in a blinded manner. Descriptive terms were used to characterize the US and mammographic appearance of the lesions from a pre-defined lexicon. A new data collection process has been implemented for the upcoming year. Final results will be deferred until the end of this year.

14. SUBJECT TERMS Mammography			15. NUMBER OF PAGES 12
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.
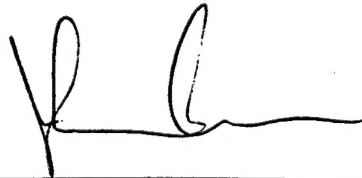
____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

_✓_ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

1/31/99

_____
PI - Signature   Date

Table of Contents

1. Introduction

Modifications to Project Duration and Aims

This was originally a two-year project scheduled for completion by Jan 31, 1999. During this past (second) year of the project, two major changes were approved by the USAMRMC, namely a change in PI and no-cost extension of the grant.

First, the original PI, Dr. Jay Baker, left this institution effective July 1, 1998. The PI title was transferred to Dr. Joseph Lo within the same department. Dr. Lo was not on the original personnel list and did not have any experience with US breast imaging, but he had collaborated extensively with Dr. Baker in developing ANNs for breast cancer diagnosis in mammography, resulting in the co-authoring of many peer-reviewed publications [1-5]. In addition to the change in PI, there were several adjustments to the budget, including adding Dr. Lo as the PI, retaining Dr. Baker as a consultant, and procuring some new equipment and resources to facilitate the transition.

Second, the transfer of PI resulted in many unexpected challenges, both bureaucratic and scientific. Consequently, it was not possible to accomplish many of the goals originally scheduled for the second year, nor were the budgeted funds expended. The new PI therefore requested and received approval for a no-cost extension for a third year, in order to accomplish the aims of the project by January 31, 2000. In this process, the USAMRMC approved a new statement of work, which is listed in the body of this document later. This new statement refocused the aims of the second year and deferred most of them to this coming third year.

Due to these changes, this document is the second annual report rather than the final report, which has been deferred to January 31, 2000. This report will summarize the achievements from the second year, and outline plans for accomplishing the remaining aims during the third and final year.

Purpose, Scope, and Background

Diagnostic imaging of the breast is dominated by the modalities of mammography and ultrasonography. Mammography is very sensitive at detecting 90% of breast cancers, but it is not very specific, resulting in a false positive biopsy rate of approximately 65% [6, 7]. These false positive biopsies result in a considerable emotional, physical, and psychological burden to the patients, as well as a significant financial burden to society.

Currently the only widely accepted role of ultrasound (US) in diagnostic breast imaging is the differentiation of simple cysts from solid breast masses [8]. One study has suggested it is possible to differentiate benign vs. malignant masses based upon US features [9], but this work is unduplicated and controversial [10]. US has considerable potential, however, because of its low cost, use of nonionizing radiation, and wide

availability. In particular, it may be useful in helping to assess masses identified by screening mammography or physical exam.

US suffers from two primary limitations. First, although some US features raise the suspicion for breast cancer, no individual US feature is specific enough to predict malignancy. Secondly, US is highly operator dependent with potential for considerable inter- and intra-observer variability in not only the images obtained by also the interpretation of those images. This study attempts to address both these limitations.

The purpose of this study is to develop artificial neural network (ANN) models to assist radiologists in interpreting US images of the breast. An ANN is a computerized predictive model capable of learning patterns from a set of training data, then generalizing its predictions robustly to similar cases it has never seen before. For example, in previous work at this institution, an ANN was developed to predict whether breast lesions were benign or malignant using mammographic features extracted by expert mammographers and patient history findings from the medical record [1].

An ANN may be a valuable tool in assisting radiologists to evaluate US images. First, it can capture subtle relationships among multiple image findings, with an accuracy and consistency matching or surpassing that of expert radiologists, as demonstrated in previous work at this institution [4, 5, 3]. Rather than evaluating the effect of a single US feature, the ANN iteratively and nonlinearly combines all the findings, thus potentially making breast US a more accurate study for diagnosing breast cancer. Second, ANNs are well suited to reduce the inter- and intra-observer variability in the interpretation of US exams. In another previous study, it was shown that ANNs can handle inter-observer inconsistencies in the input findings and still produce robust breast cancer diagnoses that were significantly more consistent than radiologists [11].

2.   Body

The original statement of work as outlined in the first year's annual report is repeated below. The revised statement of work follows, along with a description of how they differ and an overview of what was accomplished for each aim.

Original Statement of Work (1/1/1997 to 1/31/1999)

Technical Objective 1: Develop an artificial neural network (ANN) to predict biopsy outcomes from US findings.
  • Create a database of US, mammographic, and physical exam findings, as well as medical and family history data for women with solid breast masses and histologically-proven diagnoses.
  • Build a neural network from the database to predict the presence of breast cancer. Maximize the specificity while maintaining perfect or near-perfect sensitivity. Evaluate this computer-aided diagnosis system using "round robin" techniques.

Technical Objective 2: Evaluate the diagnostic accuracy of the neural network system in a clinical setting.
  • Apply the ANN to approximately 100 cases obtained after the sixth month of the project that are not included in those training cases used to develop the neural network. Determine whether the network generalizes from training cases to new test cases. Test different input features to improve the ability of the network to generalize.

Technical Objective 3: Evaluate the usefulness of the ANN in improving observer variability in US examination of breast masses.
  • Create a database of approximately 100 cases in which three radiologists read independently complete US examinations of the same solid nodules. Calculate the inter-observer variability of the radiologists' findings of breast US examination of these 100 cases. Use Cohen's kappa statistic to measure observer variability.

Revised Statement of Work for Year 3 (2/1/1999 to 1/31/2000)

Using remaining funds in the grant, we propose to accomplish the following specific aims during the extended, third year of this project:

A. Months 1-10. Resume collection of retrospective cases. We will attempt to double the current database of approximately 100 patient cases to 200 overall. For each patient, we will record ultrasound (US) and mammography findings and patient history data.

B. Months 9-10. Given the larger database of patient cases, optimize the performance of an artificial neural network (ANN) to predict malignancy among breast masses. The ability of the ANN to generalize from training cases will be evaluated using retrospective data sampling rather than prospective clinical evaluation.

C. Month 11. Evaluate the contribution of different input features in order to develop a simplified ANN that maintains diagnostic performance while requiring fewer features.

D. Month 12. Evaluate the usefulness of the ANN in improving observer variability in US examination of breast masses. Specifically, compare the consistency and accuracy of the radiologists' assessments with that of the predictions of the ANN using the radiologists' findings as inputs.

(The new tasks were renumbered as A–D in this report to minimize confusion with technical objectives 1–3 in the original statement of work.)

Overview of Progress for Each Aim

In this section we will briefly describe how each of the original technical objectives (1-3) were accomplished and/or modified into the new tasks (A–D) scheduled for the coming year.

Original objective 1 sought to develop a database and train a preliminary ANN. This was accomplished during the first year, using 65 patient cases collected by the original PI. These results were documented in detail in the first annual report and will not be repeated. The new PI was able to duplicate these results using the round robin data sampling scheme rather than cross-validation, to minimize any performance bias due to the small size of the database. These revised results were fortunately quite similar to those reported before. The new PI presented these results at the First International Workshop on Computer-Aided Diagnosis sponsored by the University of Chicago department of radiology in Chicago, IL [12].

Original objective 2 of evaluating the ANN using 100 new cases was deemed inappropriate. With such relatively few cases, arbitrarily dividing them into two parts can considerably bias the results. The ANN does not have enough cases to learn the diagnostic problem adequately, nor can its ability to generalize be assessed with so few testing cases. The revised aims address these limitations. The new task A seeks to collect 200 cases, which in task B will be used to test the ANN retrospectively using the round robin data sampling technique. (Ironically this was proposed by the original PI but not implemented.) Data sampling techniques such as the round robin rely upon many subdivisions of the data such that all cases are used for both training and testing, while still assuring independence between the two.

Collecting these new cases proved to be far more difficult than initially anticipated. The original PI acted in the dual role of radiologist and scientist. As such he could readily identify appropriate cases, namely those with US examinations which revealed a solid breast mass that underwent biopsy. He was then able to interpret these cases at his convenience and extract the required findings. During this past year he was able to collect an additional 35 cases, bringing the total to 100 cases.

The new PI is not a radiologist, however, and was thus unable to either identify or interpret these cases. After several abortive attempts, a new data collection scheme has been devised and implemented for the coming year. The breast imaging section at this institution maintains a limited database of all patients which undergo US-guided biopsies. These cases qualify for this study because they are almost all solid breast masses with US exams and all have definitive histologic outcomes. These specific cases are being presented to the new radiologist reader in the project, Dr. Mary Scott Soo, who is also the new head of breast imaging at this institution. Dr. Soo extracts both the mammographic and US findings retrospectively in routine hourly sessions to minimize fatigue. With approximately 100 qualifying cases in the latter half of 1998 alone, and with Dr. Soo committed as budgeted personnel, there appear to be no further obstacles to achieving our goal of 200 cases overall by the 10th month of the third year, in accordance with **Task A** of the new aims. During that 10th month we will re-optimize a new ANN using the 200 cases using the aforementioned data sampling technique, in accordance with **Task B** of the new aims. Finally, **Task C** of the new aims is an extension of the approach described in the original objective 2 which suggested evaluating the effect of using fixed combinations of different input features, such as (1) the 7 US findings plus patient age, (2) the 7 US findings plus the 6 mammography findings, or (3) the 7 US findings plus the 6 mammography findings and patient age. **Task C** of the new aims will additionally employ an empirical technique developed by the new PI to determine an optimal combination of the findings [3].

Original objective 3 of assessing the usefulness of the ANN in reducing observer variability was initiated during the first year, reported in the first annual report, and followed through to a peer-reviewed publication during the second year. In brief, 60 cases were read independently by 5 radiologists, and the consistency of their US findings as well as diagnostic assessment of likelihood of malignancy were measured. It was found that "considerable" variability existed for choosing terms for describing US findings as well as predicting the diagnosis. This work was accomplished by the original PI in his revised role as a consultant after his departure from this institution. This manuscript has been accepted and is currently in press for publication in 1999.

The original objective 3 suggested that separate ANNs be developed for each radiologist's US findings in order to see if the ANN can reduce the inter-observer variability in predicting malignancy. This goal will not be pursued because the initial results for those 60 cases indicate essentially perfect performance by both radiologists and ANN already, leaving no room for statistically significant improvements. During the coming year, more than tripling the size of the database to approximately 200 cases should help to provide more meaningful results in terms of ANN as well as radiologist performance. It is beyond the limited scope and resources of this project, however, to have all these cases multiply read by experienced radiologists. As such it will not be possible to assess the consistency of the intra- or inter-observer variability in either the US findings or diagnostic assessments any further than what has already been reported in the aforementioned manuscript.

Task D of the new aims will refocus and restrict this goal considerably. The accuracy of the ANN outputs vs. the diagnostic assessments by the individual radiologist reader (Dr. Baker for the first 100 cases and Dr. Soo for the latter 100 cases) will be compared. Contrary to the stated aim of Task D, consistency will *not* be evaluated nor will multiple radiologists be used. With the larger database, it is anticipated that accuracies will no longer be near perfect, and thus more meaningful statistical comparisons may be made.

3.　　Conclusions

This past, second year has been a time of transition for this project. The transfer of PI resulted in many unexpected delays, culminating in a no-cost extension request which was approved by the USAMRMC. As a result, the results from the second year are fairly limited. Two aims from the first year were followed through to a conference proceeding and peer-reviewed manuscript. The data collection process, which is crucial to the final phase of the project, continued briefly under the charge of the former PI, resulting in 100 cases overall. This data collection process has been completely revamped under the new PI and new head of breast imaging.

Under the terms of the no-cost extension, a new statement of work was submitted to and approved by the USAMRMC. This new statement postpones many of the goals of the second year into the coming year. These goals were briefly summarized in the previous sections, with plans for achieving them explained. It is anticipated that there will be no further bureaucratic or scientific difficulties associated with this project, such that more interesting scientific results may be presented in the third, final report.

4. References

1. Baker JA, Kornguth PJ, Lo JY, Williford ME, and Floyd CE, Jr. Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. Radiology 1995; 196: 817-822.

2. Baker JA, Kornguth PJ, Lo JY, and Floyd CE, Jr. Artificial neural network: improving the quality of breast biopsy recommendations. Radiology 1995; 198: 131-135.

3. Lo JY, Baker JA, Kornguth PJ, and Floyd CE, Jr. Computer-aided diagnosis of breast cancer: artificial neural network approach for optimized merging of mammographic features. Academic Radiology 1995; 2: 841-850.

4. Lo JY, Baker JA, Frederick ED, Kornguth PJ, and Floyd CE, Jr. Predicting breast lesion malignancy and invasion using the BI-RADS mammography lexicon. Radiology 1997; 205(P): 447.

5. Lo JY, Baker JA, Kornguth PJ, and Floyd CE, Jr. Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks. Academic Radiology 1999; in press.

6. Kopans DB. The positive predictive value of mammography. AJR. American Journal of Roentgenology 1992; 158: 521-526.

7. Knutzen AM and Gisvold JJ. Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions. Mayo Clinic Proceedings 1993; 68: 454-460.

8. Jackson VP. Management of solid breast nodules: what is the role of sonography? Radiology 1995; 196: 14-15.

9. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker S, and Sisney G. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology 1995; 196: 123-134.

10. Jackson VP. Sonography of Malignant Breast Disease. Seminars in Ultrasound, CT, and MR 1989; 10: 119-131.

11. Baker JA, Kornguth PJ, and Floyd CE, Jr. Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description. AJR. American Journal of Roentgenology 1996; in press.

12. Lo JY and Floyd CE, Jr. *Computer-aided diagnosis of breast cancer*. in *First International Workshop on Computer-Aided Diagnosis*. 1998. Univ. of Chicago, Chicago, IL: Elsevier Science.